Title

The Information Theoretic Foundations of a Probabilistic and Predictive Micro and Macro Economics

Author

Judge, George

Publication Date

2012-04-20

University of California, Berkeley
Department of Agricultural &
Resource Economics

CUDARE Working Papers

Year 2012 *Paper 1123*

The Information Theoretic Foundations of a Probabilistic and Predictive Micro and Macro Economics

George Judge

The Information Theoretic Foundations of a Probabilistic and Predictive Micro and Macro Economics

George Judge*

Abstract

Despite the productive efforts of economists, the disequilibrium nature of the economic system and imprecise predictions persist. One reason for this outcome is that traditional econometric models and estimation and inference methods cannot provide the necessary quantitative information for the causal influence-dynamic micro and macro questions we need to ask given the noisy indirect effects data we use. To move economics in the direction of a probabilistic and causal based predictive science, in this paper information theoretic estimation and inference methods are suggested as a basis for understanding and making predictions about dynamic micro and macro economic processes and systems.

Key words: Information theoretic methods, State space models, First order Markov processes, Inverse problems, Dynamic economic systems.

JEL Classification: C40, C51

---

*George G. Judge, Professor in the Graduate School, 207 Giannini Hall, University of California, Berkeley, Berkeley, CA, 94720 (e-mail: gjudge@berkeley.edu).

It is a pleasure to acknowledge the helpful comments of Mike George, J. Geweke, A. Plastino, J. Stiglitz, S. Turnovsky, S. Berto-Villas and A. Zaman. The information theoretic models presented represents joint work over the years with A. Golan, M. Grendar, D. Miller and R. Mittelhammer.

1. Introduction

In a recent article, "Rethinking Macroeconomics: What Failed and How to Repair It", Stiglitz (2011) provides a multi-faceted, critical critique of the reigning paradigm in macroeconomics. The usual methodological problems are identified and suggestions are made as to how to patch things up and set things right. Certainly, a reform of the macro standard form is welcome event. In reading the Stiglitz article I had the feeling that it was not only important to rethink macroeconomics, but to also rethink all of the MME components of economic information recovery.

In the golden age of economic innovation in the 1940's, there was the feeling that micro, macro and econometrics acted as a team and that they joined forces to understand and consider solutions to economic problems. Since understanding and predicting the state of a dynamic system is essentially a statistical undertaking, in one sense it all came together in econometrics. There was a feeling of a common goal of the three MME economic foundation stones and how they might be used in a decision context and for policy purposes.

Today, there are all kinds of specialties and a major economics department may have, in any one week, as many as fifteen specialized seminars, with each member of the audience glued to their iPhone or iPad and thinking about, and working on, their specialized topic. Over time, the knowledge slices in economics have become so thin that the sum of the parts does not serve as a basis for understanding and making predictions concerning the whole. The academic pursuit of economic self-interests does not lead, as if by an invisible hand, to an understanding of economic processes and systems and has limited our ability, in a probability context, to make viable economic predictions. Although, much of the macro ills deal, among other things, with a lack of understanding of the micro foundations of macroeconomics and the single equilibrium concept, the focus in this discussion paper is on econometrics. No matter how good the macro-micro conceptual models are, their usefulness is impaired by the lack of reliable quantitative-econometric information.

1.1 The Statistical Complexity Issue

Economic-behavioral processes involve the rich ingredients of uncertainty, complexity, volatility and ambiguity. The statistical complexity econometric measure that results, reflects the unknown initial condition and the dynamics of the economic system and is a function of two ingredients: i) the choice of the probability metric space and, ii) the best distance (disequilibrium) choice in the probability space. Working models of the economy such as dynamic stochastic general equilibrium models, have little chance of success, because economic-behavioral processes and systems are seldom if ever in equilibrium and from the econometric side the estimation methods are not appropriate for solving the resulting stochastic inverse problem. Although, there is only one sampling distribution consistent with an economic system in equilibrium, there are a large number of possible ways an economic process-system may be out of equilibrium. For many econometric problems the natural solution is not a fixed distribution, but a well defined set of distributions, each with its own probability. Uncertainty about existing conditions and the dynamics of the process creates problems for model specification and prediction. Although economic relationships are, in general, simple in nature, the underlying dynamics is complicated and not well understood. This means the economic-econometric model that we use to confront the data is in some way incorrectly specified

In terms of data associated with information transfer in a dynamical system, this usually consists of mutual indirect noisy effects observations, and thus does not contain dynamic or directional information. Even introducing a lag in the mutual observations fails to distinguish information that is actually exchanged from shared information, and does not support predictions such as causal influence. Also since the model is conceptual in nature, it usually contains certain inadequacies regarding the specification of moment conditions- estimating equations that are directly connected to the data. This means the indirect noisy effects observations that are used in an attempt to identify the underlying dynamic system and to measure causal influence, involve the solution of a stochastic inverse problem, where usually the number of measurements-data points are smaller than the number of unknown parameters to be estimated. Thus this stochastic ill posed underdetermined problem cannot be solved by traditional estimation and inference methods. As a result, traditional parametric structural estimation and inference methods are fragile under this type of model and data uncertainty and are, in general, not applicable for the causal influence dynamic macro economic questions that we need to ask and the data we must use.

1.2 Coping With the Uncertainty of Uncertainty

Given these basic critical propositions, as they relate to micro and macro economic information recovery, the next question is what can we do about it? A natural solution would seem to be to use estimation and inference methods that are designed to deal with observational data and solve this type of ill posed inverse problem. In this context the Cressie-Read (CR- 1983,1988) family of likelihood functionals permits the researcher to exploit the statistical machinery of information theory to gain insights relative to the underlying causal behavior of a dynamic process that may or may not be in equilibrium. In developing this type of information theoretic econometric approach to estimation and inference the Cressie-Read single parameter family of informational functionals-divergences represent possible likelihood functions associated with the underlying sampling distribution. Information functional-likelihood-divergences of this type have an intuitive interpretation reflecting the uncertainty of uncertainty as it relates to out of equilibrium processes. This gives new meaning to what is a likelihood function and what is the appropriate way to represent the possible underlying sampling distribution of an econometric-statistical model. One possibility for implementing this approach would be to use estimating equations-moment conditions as a link to the data, and discrete members of the Cressie-Read family to identify the weighting of the possible density-likelihood functions. The outcome would reflect in a probability sense, what we know about the unknown parameters and a possible density function. The result may be a canonical hyper distribution density of possible underlying distributions. An advantage of this approach, in addition to its optimality base, is that it permits the possibility of non Gaussian like and possibly volatile distributions. Within this context the importance of developing probabilistic and predictive results from a sample of indirect noisy effects data, guides the information theoretic framework sketched ahead.

2. Minimum Power Divergence

In identifying measures that may be used as a basis for characterizing the data sampling process of indirect noisy observed data outcomes, we begin with the family of divergence measures proposed by Cressie and Read (1984) and Read and Cressie (1988). Cressie and Read (CR) developed a family of goodness-of-fit test statistics and proposed the following power divergence family of measures:

$$I(\mathbf{p},\mathbf{q},\gamma)=\frac{1}{\gamma(\gamma+1)}\sum_{i=1}^{n}p_i\left[\left(\frac{p_i}{q_i}\right)^{\gamma}-1\right]. \tag{2.1}$$

In (2.1), γ is a parameter that indexes members of the CR family, p_i's represent the subject probability distribution and the q_i's, are interpreted as reference probabilities. Being probabilities, the usual probability distribution characteristics of $p_i, q_i \in [0,1] \forall i, \sum_{i=1}^{1} p_i = 1$, and $\sum_{i=1}^{n} q_i = 1$ are assumed to hold.

The CR family of power divergences is defined through a class of additive convex functions and the CR power divergence measure encompasses a broad family of test statistics, and leads to *a broad family of likelihood functions* within a moments-based estimation context. In addition the CR measure exhibits proper convexity in **p**, for all values of γ and **q**, and embodies the required probability system characteristics, such as additivity and invariance with respect to a monotonic transformation of the divergence measures. In the context of extremum metrics, the general Cressie-Read (1984) family of power divergence statistics represents a flexible family of pseudo-distance measures from which to derive empirical micro and macro probabilities.

The CR statistic is a one parameter family of divergence measures that indexes a set of empirical goodness-of-fit and estimation criteria. As γ varies, the resulting estimators that minimize power divergence exhibit qualitatively different sampling behavior. Using empirical sample moments such as $h(Y, X, Z; \beta) = n^{-1}[Z'(Y - X\beta) = 0$, as constraints, a solution to the stochastic inverse problem, based on the optimized value of $I(\mathbf{p}, \mathbf{q}, \gamma)$, is one basis for representing a range of data sampling processes and likelihood function values.[1]

[1] To place the CR family of power divergence statistics in an entropy perspective, we note that there are corresponding Renyi (1961, 1970) and Tsallis (1988) families of entropy functionals-divergence measures. As demonstrated by Gorban, Gorban and Judge (2010), over defined ranges of the divergence measures, the CR and entropy families are equivalent. Relative to Renyi and Tsallis, the CR family has a more convenient normalization factor 1/(□(□+1)), and has proper convexity for all powers, both positive and negative. The CR family has the separation of variables for independent subsystems (Gorban, et al.,2010) over the range of lambda. This separation of variables permits the partitioning of the state space and is valid for divergences in the form of a convex function.

2.1 Minimum Power Divergence Estimation

In a linear econometric model context, if we use (2.1) as the goodness-of-fit criterion, along with moment-estimating function information, the estimation problem based on the CR divergence measure (CRDM) can, for any given choice of the γ parameter, be formulated as the following extremum-type estimator for $\boldsymbol{\beta}$**:**

$$\hat{\boldsymbol{\beta}}(\gamma) = \arg\min_{\boldsymbol{\beta}\in\mathbf{B}}\left[\min_{\mathbf{p}}\left\{ I(\mathbf{p},\mathbf{q},\gamma) \mid \sum_{i=1}^{n} p_i \mathbf{z}'_{i.}(y_i - \mathbf{x}_{i.}\boldsymbol{\beta}) = \mathbf{0}, \sum_{i=1}^{n} p_i = 1, p_i \geq 0 \, \forall i \right\}\right] \quad (2.2)$$

where, $\mathbf{q}$ is usually taken as a uniform distribution. This class of estimation procedures is referred to as *Minimum Power Divergence (MPD)* estimation (see Judge and Mittelhammer, 2011).

It is important to note that the *family* of power divergence statistics, that are defined by (2.1), is symmetric in the choice of which set of probabilities are considered as the subject and reference distribution arguments of the function (2.4). In particular, whether the statistic is designated as $I(\mathbf{p},\mathbf{q},\gamma)$ or $I(\mathbf{q},\mathbf{p},\gamma)$ *the same collection* of members of the family of divergence measures, are ultimately spanned.

Two discrete CR divergences for $\mathrm{I}(\mathbf{p},\mathbf{q},\gamma)$ have received the most attention in the literature. We utilize the abbreviated notation, $\mathrm{CR}(\gamma) \equiv \mathrm{I}(\mathbf{p},\mathbf{q},\gamma)$, where the arguments $\mathbf{p}$ and $\mathbf{q}$ are tacitly understood to be evaluated at relevant vector values. In the two special cases where, $\gamma = 0$ or -1, , the notations $\mathrm{CR}(0)$ and $\mathrm{CR}(-1)$ are to be interpreted as the continuous limits, $\lim_{\gamma\to 0}\mathrm{CR}(\gamma)$ and $\lim_{\gamma\to -1}\mathrm{CR}(\gamma)$, respectively.

Minimizing $CR(-1)$ leads to the traditional maximum empirical log-likelihood (MEL) objective function, and to the Kullback-Leibler (KL)(1951) and Kullback(1959) divergence $D_{CR,-1} = n^{-1}\sum_{i=1}^{n}\ln(p_i)$. The specification $CR(0)$ leads, under a uniform reference distribution, to the objective function $D_{CR,0} = -\sum_{i=1}^{n} p_i \ln(p_i)$ and this divergence is equivalent to Shannon's (1948) relative entropy measure.

In regard to inference with the MPD (CR(γ) family) estimators, under the usual assumed regularity conditions, all of the MPD estimators of $\boldsymbol{\beta}$ are consistent and asymptotically normally distributed. They are also asymptotically efficient, relative to the optimal estimating function (OptEF) estimator (Baggerly, 1998), when a uniform distribution, or equivalently the empirical distribution function (EDF), is used for the reference distribution. The solution to the constrained optimization problem yields optimal estimates, $\hat{\mathbf{p}}(\gamma)$ and $\hat{\boldsymbol{\beta}}(\gamma)$, that cannot, in general, be expressed in closed form, and thus must be obtained using numerical methods.

Since the likelihood function and the sample space are inexplicably linked, given a sample of indirect noisy observations and corresponding moment conditions, it would be useful to have an optimum choice of a member of the CR family. Usually in traditional econometrics, given a sample of data and corresponding moment conditions, there is ambiguity-uncertainty regarding the choice of likelihood functions. At this point we again emphasize that in an economic disequilibrium situation the solution may not be a fixed distribution, but a well defined set of distributions- likelihood functions-PDFs.

2.2 Identifying the Probability Space

Given the CR family of divergence measures (2.1), indirect noisy dynamic data and linear functionals in the form of moments, the next question is how to go about identifying the underlying probability distribution function-probability space of a system or process that may or may not be in equilibrium. Divergence measures permit us to exploit the statistical machinery of information theory to gain an insight into the PDF behavior of dynamic disequilibrium economic systems and processes. The likelihood functionals-PDFs-divergences, have an intuitive interpretation in terms of uncertainty and measures of distance. Many formulations have been proposed for a proper selection of the probability space, but their applicability depends on characteristics of the data, such as stationarity and the noise process. In the section ahead we suggest how to make use of the CR family of divergence measures to choose the optimal probability system under quadratic or Kullback-Leibler loss.

In Section 2.1, we used the CR power divergence measure (2.1), to define a family of likelihood functions. Given this family of likelihood functions, one might, as in (2.3), consider a parametric family of concave entropy-likelihood functions, which satisfy additivity and trace conditions. Given a family of divergence functions,one might follow Gorban (1984) and Gorban and Karlin (2003) and consider a parametric family of convex information divergences, which satisfy additivity and trace conditions. Convex combinations of CR(0)) and (CR-1) produce a remarkable family of divergences-distributions. Using the CR divergence measures, this parametric family is essentially the linear convex combination of the cases where $\gamma = 0$ and $\gamma = -1$. This family is tractable analytically and provides a basis for joining (combining) statistically independent subsystems. When the base measure of the reference distribution **q** is taken to be a uniform probability density function, we arrive at a one-parameter family of additive convex dynamic Lyapunov functions In this context, one would be effectively considering the convex combination of the MEL and maximum empirical exponential likelihood (MEEL) measures. From the standpoint of extremum-minimization with respect to **p**, the generalized divergence family , under uniform q, reduces to

$$S_{\alpha}^{*} = \sum_{i=1}^{n} \left(-(1-\alpha)\, p_i \ln(p_i) + \alpha \ln(p_i) \right),\ \ 0 \le \alpha \le 1. \tag{2.3}$$

In the limit, as $\alpha \to 0$, the Kullback-Leibler or minimum I divergence I(**p** || **q**) of the probability mass function **p,** with respect to **q**, is recovered. As $\alpha \to 1$, the MEL stochastic inverse problem I(**q||p)** results. *This generalized family of divergence measures permits a broadening of the canonical distribution functions and provides a framework for developing a loss-minimizing estimation rule(Hall,1987, Jeffreys, 1983).* In line with the complex nature of the problem, in the section to follow, we demonstrate convex estimation rules, that choose among MPD-type estimators to minimize quadratic risk (QR).

2.3 A Minimum Quadratic Risk (QR) Estimation Rule

In this section, we use the well-known squared error-quadratic loss criterion and associated QR function to choose among a given set of discrete alternatives for the CR goodness-of-fit measures and associated estimators for $\boldsymbol{\beta}$. The method seeks to define the convex combination of a set of estimators for $\boldsymbol{\beta}$ that minimizes QR, where each estimator is defined by the solution to the extremum problem

$$\hat{\boldsymbol{\beta}}(\gamma) = \arg\max_{\boldsymbol{\beta}\in\mathbf{B}} \left[\max_{\mathbf{p}} \left\{ -I(\mathbf{p},\mathbf{q},\gamma) \mid \sum_{i=1}^{n} p_i \mathbf{Z}_{i.}' \left(Y_i - \mathbf{X}_{i.}\boldsymbol{\beta} \right) = \mathbf{0}, \sum_{i=1}^{n} p_i = 1, p_i \geq 0 \; \forall i \right\} \right]. \quad (2.4)$$

The squared error loss function is defined by $\ell\left(\hat{\boldsymbol{\beta}},\boldsymbol{\beta}\right) = \left(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}\right)'\left(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}\right)$ and has the corresponding QR function given by

$$\rho\left(\hat{\boldsymbol{\beta}},\boldsymbol{\beta}\right) = E\left[\ell\left(\hat{\boldsymbol{\beta}},\boldsymbol{\beta}\right)\right] = E\left[\left(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}\right)'\left(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}\right)\right]. \quad (2.5)$$

The convex combination of estimators is defined by

$$\overline{\boldsymbol{\beta}}(\boldsymbol{\alpha}) = \sum_{j=1}^{J} \alpha_j \hat{\boldsymbol{\beta}}(\gamma_j), \; where \;\; \alpha_j \geq 0 \; \forall j, \; and \; \sum_{j=1}^{J} \alpha_j = 1. \quad (2.6)$$

The optimum use of the discrete alternatives under QR is determined by choosing the particular convex combination of the estimators that minimizes QR, as

$$\overline{\boldsymbol{\beta}}(\hat{\boldsymbol{\alpha}}) = \sum_{j=1}^{J} \hat{\alpha}_j \hat{\boldsymbol{\beta}}(\gamma_j), \text{ where } \hat{\boldsymbol{\alpha}} = \arg\min_{\boldsymbol{\alpha}\in CH} \left\{ \rho\left(\overline{\boldsymbol{\beta}}(\boldsymbol{\alpha}),\boldsymbol{\beta}\right) \right\} \quad (2.7)$$

and *CH* denotes the *J*-dimensional convex hull of possibilities for the $J \times 1$ $\boldsymbol{\alpha}$ vector, defined by the nonnegativity and adding-up conditions. This convex combination rule represents one possible method of choosing the unknown gamma parameter. The CR family of likelihood functionals has an intuitive interpretation in terms of uncertainty measures and divergence between probability distributions. This permits discrimination in the context of density estimation and permits us to gain insights into the probability distribution-PDF behavior of dynamic state space economic processes and, importantly, to make predictions. Finally we note that i) information theoretic methods permit us to go outside the usual density specification and consider non Gaussian distributions, that in some cases may reflect volatility (bubbles) associated with the economic process, and ii) predictability in dynamic economic models appears naturally using information theoretic functions. The static probabilities in mutual observations can be given a directional meaning when transition probabilities are introduced.

2.4 A Dynamic Probabilistic Information Recovery Process

To incorporate directional dynamic structure we us use transition probabilities, and focus on first-order Markov chain models of events with a finite number of outcomes measured at discrete time intervals (Lee, Judge and Zellner, 1977, Miller, 2007, Miller and Judge, 2012 and Kristensen and Shin, 2012). From the micro perspective, the decision outcomes for agent $i = 1, 2, \ldots, n$ are denoted $Y(i,k,t)$ with finite states $k=1,2,\ldots,K$ at time $t=0,1,2,\ldots T$. If the decision outcomes exhibit first-order Markov character, the dynamic behavior of the agents may be represented by conditional transition probabilities $p\left(j,k,t\right)$, which represent the probability that agent *i* moves from state *j*=1, 2,…,*K*, to state *k* at a time *t*. Given observations on the micro behavior *Y*(*i*,*k*,*t*), the discrete Markov decision process framework may be used to model the agent-specific dynamic economic behavior. In this paper the focus is on conditional Markov models in which $p(j,k,t)$ varies with *t*. In the conditional case, we have $T\times(K-1)\times(K-1)$ unknown transition probabilities in the $(K-1)\times(T-1)$ estimating equations. Thus the estimation problem is ill posed and traditional estimation methods are not applicable.

Since the available data may be partial or incomplete, one key step is linking the sample analog of the Markov process to the indirect noisy observations

$$Y(k,t)=\sum_{j=1}^{K}Y(j,t-1)\,p(j,k,t)+e(k,t) \tag{2.8}$$

For empirical purposes, the remaining task is to choose a feasible specification of the statistical model of the Markov transition probabilities.

This new class of conditional Markov models is based on a set of estimating equations-moment equations

$$E\left[z_t'e(k,t)\right]=0 \tag{2.9}$$

where $\mathbf{z}_t$ is an appropriate set of instrumental-intervention variables. By substitution of (4.1) into (4.2), we form a set of estimating equations that are expressed in terms of the unknown transition probabilities. Given there may be many feasible transition probability models that satisfy the moment-estimating equations, the next step is to provide a model for the data and a basis for identifying parametric data sampling distributions and likelihood functions in the form of distances in probability space. In this context, again considerthe Cressie-Read (Cressie and Read, 1984; Read and Cressie, 1988; Baggerly, 1998; Judge and Mittelhammer, 2011), family of power divergence measures that may be defined for a set of first-order finite and discrete conditional Markov probabilities as

$$I\left(p,q,\alpha\right)=\frac{2}{\alpha\left(1+\alpha\right)}\sum_{t=1}^{T}\sum_{j=1}^{K}\sum_{k=1}^{K}p\left(j,k,t\right)\left[\left(\frac{p\left(j,k,t\right)}{q\left(j,k,t\right)}\right)^{\alpha}-1\right] \tag{2.10}$$

provides access to a rich set of distribution functions that encompasses a family of estimation objective functions indexed by discrete probability distributions convex in p.

In order not to introduce subjective information, the reference distribution will be specified as discrete uniform distributions. Formally, the MPD problem may be solved by choosing transition probabilities *p* to minimize $I\left(p,q,\alpha\right)$ (for some α), subject to the sample analogs of (3.2)

$$\sum_{t=1}^{T}\mathbf{z}_t'\left(Y\left(k,t\right)-\sum_{j=1}^{K}Y\left(j,t-1\right)p\left(j,k,t\right)\right)=\mathbf{0} \tag{2.11}$$

for each j = 2, ..., K and the row-sum constraint

$$\sum_{k=1}^{K}p\left(j,k,t\right)=1 \tag{2.12}$$

for all *j* and *t*. When $\alpha\rightarrow 0$ the entropy-likelihood functional is

$$-I\left(p,q,\alpha\rightarrow 0\right)\propto-\sum_{t=1}^{T}\sum_{j=1}^{K}\sum_{k=1}^{K}p\left(j,k,t\right)\ln\left(p\left(j,k,t\right)\right) \tag{2.13}$$

and has a general logistic functional form. When $\alpha \to -1$, the entropy-empirical likelihood functional (Owen, 2001) is

$$-I\left(p,q,\alpha \to -1\right) \propto \sum_{t=1}^{T}\sum_{j=1}^{K}\sum_{k=1}^{K} \ln\left(p\left(j,k,t\right)\right). \tag{2.14}$$

Since the likelihood function and the sample space for the MPD estimation problem are inexplicably linked, it would be useful, given a sample of indirect noisy observations and corresponding moment conditions, to have, as in Section 3, an optimum choice of a member of the CR family.

If we consider a parametric family of concave entropy-likelihood functions, which satisfy additivity and trace conditions this parametric family is essentially the linear convex combination of the cases where $\alpha \to 0$ and $\alpha \to -1$.From the standpoint of extremum-minimization with respect to the transition probabilities, the generalized divergence family reduces to

$$S_{\beta}^{*} = -\left(1-\beta\right)\sum_{t=1}^{T}\sum_{j=1}^{K}\sum_{k=1}^{K} p\left(j,k,t\right)\ln\left(p\left(j,k,t\right)\right) + \beta\sum_{t=1}^{T}\sum_{j=1}^{K}\sum_{k=1}^{K}\ln\left(p\left(j,k,t\right)\right) \tag{2.15}$$

a convex combination of the CR family members in (4.7) and (4.9). *As noted in Section 3 this generalized family of divergence measures permits a broadening of the canonical distribution functions and provides a framework for developing a loss-minimizing estimation rule.*

3. Summing Up

Economics seeks, given an information base, to provide a general theory of choice. From an economic policy standpoint there is much to be unhappy about concerning the disequilibrium nature of our economy and the uncertain nature of the predictive choices. Although, much of this unhappiness seems to be focused on the macro area of economics, in this paper the focus has been on quantitative economic information and the econometric component. Since the 1940s we have survived the Klein-MIT-Fed type large scale econometric models, Vector Auto Regressions-very awful regressions, rational expectations, and the arbitrariness of calibration that essentially picks a few parameters at random which may match a few arbitrarily chosen moments or empirical regularities. We have noted that although the desire for information has been on causal influence-information recovery, the intersection of this objective and the indirect noisy effects data and traditional direct econometric methods that are supposed to provide this type of information, is in many cases an empty set. The data for the most part are observational, and the estimation-information recovery component requires methods designed for solving stochastic ill posed inverse problems.

As a solution, we have suggested the use of information theoretic econometric methods that are designed to cope with these types of inverse problems and provide policy choices that are not drawings from a uniform-maximum entropy distribution. The message of this paper is that unless we develop and apply econometric models and methods that are appropriate to the data and economic problems-questions at hand, our ability to understand and recover empirical system probabilities and to provide accurate predictions about dynamic economic processes, is going to continue to be uninformed and limited. Furthermore, our brightest and best graduate students are going to continue to be taught econometric history.

4. References

Baggerly, K., 1998, Empirical Likelihood as a Goodness of Fit Measure. Biometrika 85, 535–47.

Cressie, N. and T. Read, 1984, Multinomial Goodness–of–Fit Tests., Journal of the Royal Statistical Society, Series B 46, 440–64.

Gorban, A.N., 1984, Equilibrium Encircling: Equations of Chemical Kinetics and Their Thermodynamic Analysis, Nauka: Novosibirsk.

Gorban, A.N. and I.V. Karlin, 2003, Family of Additive Entropy Functions out of Thermodynamic Limit. Physical Review E 67, 016104.

Gorban, A.N., P.A. Gorban and G. Judge, 2010, Entropy: The Markov Ordering Approach. Entropy 12, 1145-1193.

Hall, P. 1987, On Kullback-Liebler Loss and density estimation, The Annals of Statistics 15, 1491-1519.

Jeffreys, H., 1946, An Invariant form for the prior probability in Estimation Problems. Proceedings of the Royal Society A 186, 453-461.

Jeffreys, H., 1983, Theory of Probability, Clarendon Press, Oxford, 3rd ed.

Judge, G, and R. Mittelhammer, 2011, An Information Theoretic Approach to Econometrics, Cambridge University Press.

Kullback, S., 1959, Information Theory and Statistics, New York: John Wiley and Sons.

----- and R. Leibler, 1951, On Information and Sufficiency. Annals of Mathematical Statistics 52, 79–86.

Kristensen, D. and Y. Shin, 2012, Estimation of Dynamic Models With Nonparametric simulated Maximum Likelihood. Journal of Econometrics 167, 76-94.

Lee, T. C., G. Judge, and A. Zellner, 1977, Estimating the Parameters of the Markov Probability Model from Aggregate Time Series Data, 2nd ed., North–Holland.

Miller, D., 2007, Behavioral Foundations For Conditional Markov Models For Aggregate Data. Working Paper, University of Missouri.

Miller, D and G, Judge, 2012, Information Recovery in a Dynamic Statistical Model", Working paper, UC, Berkeley.

Owen, A., 2001, Empirical Likelihood, Boca Raton, FL: Chapman and Hall.

Read, T. and N. Cressie, 1988, Goodness–of–Fit Statistics for Discrete Multivariate Data, New York: Springer–Verlag.

Renyi, A. , 1961. On measures of entropy and information. In Proceedings of the 4th Berkeley Symposium on Mathematics, Statistics and Probability 1960, Berkeley, CA, USA, 20 June-30 July 1960; University of California Press: Berkeley, CA, USA, 1961; Volume 1, pp. 547-561.

Renyi, A., 1970. Probability Theory; North-Holland: Amsterdam, The Netherlands.

Shannon, C.E., 1948. A mathematical theory of communication. Bell Syst. Tech. J., 27, 379-423.

Stiglitz, J., 2011, Rethinking Macroeconomics: What Failed and How to Repair It. Journal of the European Economics Association 9, 591-645.

Tsallis, C., 1988. Possible generalization of Boltzmann-Gibbs statistics. J. Stat. Phys., 52, 479-487.